Open access • Proceedings Article • DOI:10.5244/C.16.19

Alignment using Spectral Clusters — Source link

Marco Carcassoni, Edwin R. Hancock

Institutions: University of York

Published on: 01 Jan 2002 - British Machine Vision Conference

Topics: Adjacency matrix, Graph energy, Adjacency list, Spectral graph theory and Laplacian matrix

Alignment using Spectral Clusters

Marco Carcassoni and Edwin R. Hancock
Department of Computer Science, University of York,
York, Y01 5DD, UK.
email: {marco,erh}@cs.york.ac.uk

Abstract

This paper describes a hierarchical spectral method for the correspondence matching of point-sets. Conventional spectral methods for correspondence matching are notoriously susceptible to differences in the relational structure of the point-sets under consideration. In this paper we demonstrate how the method can be rendered robust to structural differences by adopting a hierarchical approach. We show how the point-clusters associated with the most significant spectral modes can be used to locate correspondences when significant contamination is present.

1 Introduction

Spectral graph theory is a term applied to a family of techniques that aim to characterise the global structural properties of graphs using the eigenvalues and eigenvectors of the adjacency matrix [1]. Although the subject has found widespread use in a number of areas including structural chemistry and routeing theory, there have been relatively few applications in the computer vision literature. The reason for this is that although elegant, spectral graph representations are notoriously susceptible to the effect of structural error. In other words, spectral graph theory can furnish very efficient methods for characterising exact relational structures, but soon breaks down when there are spurious nodes and edges in the graphs under study.

There are several concrete examples in the pattern analysis literature. Umeyama has an eigendecomposition method that recovers the permutation matrix that maximises the correlation or overlap of the adjacency matrices for graphs of the same size [13]. Horaud and Sossa [5] have adopted a purely structural approach to the recognition of line-drawings. Their representation is based on the immanantal polynomials for the Laplacian matrix of the line-connectivity graph. By comparing the coefficients of the polynomials, they are able to index into a large data-base of line-drawings. Shapiro and Brady [11] have developed a method which draws on a representation which uses weighted edges. They commence from a weighted adjacency matrix (or proximity matrix) which is obtained using a Gaussian function of the distances between pairs of points. The eigen-vectors of the adjacency matrix can be viewed as the basis vectors of an orthogonal transformation on the original point identities. In other words, the components of the eigenvectors represent mixing angles for the transformed points. Matching between different point-sets is effected by comparing the pattern of eigenvectors in different images. Finally, a number of authors have used spectral methods to perform pairwise clustering on image data. Shi and Malik [12] use the second eigenvalue to segment grey-scale images by performing an eigen-decomposition on a matrix of pairwise attribute differences. Inoue and Urahama [6] have shown how the sequential extraction of eigen-modes can be used to cluster pairwise

BMVC 2002 doi:10.5244/C.16.19

pixel data. Sengupta and Boyer [9] have used similar ideas to find significant perceptual arrangements of line-segments.

The focus of this paper is the use of property matrix spectra for correspondence matching. As mentioned above, spectral methods offer an attractive route to correspondence matching since they provide a compact and easily computed representation that can be used to characterise graph structure at the global level. If used effectively, the spectral representation can be used for rapid matching by comparing patterns of eigenvalues or eigenvectors. However, their shortcoming is their fragility to the addition of noise and clutter. For instance, although the methods of Umeyama [13], Horaud and Sossa [5] and Shapiro and Brady [11] work well for graphs that are free of structural contamination, they do not work well when the graphs are of different size.

Our aim in this paper is to consider how spectral methods can be rendered robust for the correspondence matching of point-sets which contain significant structural difference. To do this we adopt a hierarchical approach. We observe that the modes of the proximity matrix can be viewed as pairwise clusters. Further, we note that although the coefficients of the modal matrix may be unstable under the structural modification of the proximity matrix, the physical location of the clusters will be less sensitive. We exploit these two observations to develop a hierarchical matching method. We commence by identifying the most significant modes of the proximity matrix, i.e. the largest clusters. We then compute the physical locations of the cluster-centres and compute the associated proximity matrix. Finally, we match by performing spectral analysis on the cluster-centre proximity matrix.

2 Point Correspondence

The spectral approach to point correspondence introduced by Shapiro and Brady [11] commences by enumerating a point proximity matrix. This is a continuous or weighted counterpart of the graph adjacency matrix. Rather than setting the elements to unity or zero depending on whether or not there is a connecting edge between a pair of nodes, the elements of the proximity matrix are weights that reflect the strength of a pairwise adjacency relation. The weights of the proximity matrix are computed by taking a Gaussian function of the interpoint distances, Once the proximity matrix is to hand, then correspondences are located by computing its eigenvectors. The eigenvectors of the proximity matrix become the columns of a transformation matrix which operates on the original point identities. The rows of the transformation matrix represent the components of the original points in the directions of the eigenvectors. We can locate point correspondences by searching for rows of the transformation matrix which have maximal similarity.

Unfortunately there are two drawbacks with this spectral method of correspondence. Firstly, there is no clear reason to use Gaussian weighting in favour of possible alternatives. Moreover, the Gaussian weighting may not be the most suitable choice to control the effects of pattern distortion due to point movement under measurement error or deformation under affine or perspective geometry. Secondly, the method proves fragile to structural differences introduced by the addition of clutter or point drop-out. In a recent paper we have addressed the first of these problems by using robust error kernels to computer the proximity matrix [2]. Here we focus on the second problem, and develop a hierarchical method matching point-sets.

2.1 Prerequisites

We are interested in finding the the correspondences between two point-sets, a model point-set $\mathbf{z}$ and a data point-set $\mathbf{w}$. Each point in the image data set is represented by an

augmented position vector of homogeneous co-ordinates $\vec{w}_i = (x_i, y_i, 1)^T$ where i is the point index. We will assume that all these points lie on a single plane in the image. In the interests of brevity we will denote the entire set of image points by $\mathbf{w} = \{\vec{w}_i, \forall i \in \mathcal{D}\}$ where $\mathcal{D}$ is the point set. The corresponding fiducial points constituting the model are similarly represented by $\mathbf{z} = \{\vec{z}_j, \forall j \in \mathcal{M}\}$ where $\mathcal{M}$ denotes the index-set for the model feature-points $\vec{z}_j$. We use the binary indicator $s_{i,j}$ to indicate the state of correspondence between the data-points and the model-points. If $s_{i,j} = 1$, then the data-point with co-ordinate vector $\vec{w}_i$ is in correspondence with the model-point with co-ordinate vector $\vec{z}_j$.

2.2 Point Proximity matrix

The role of the weighting function used to compute the elements of the proximity matrix is to model the probability of adjacency relations between points. In Shapiro and Brady's original work the weighting function was the Gaussian [11]. However, we have recently shown that alternative weighting functions suggested by the robust statistics literature offer significant improvements [2].

According to robust statistics, the effects of outliers can be controlled by weighting according to the error-residual. Suppose that $\Gamma_s(\eta)$ is a weighting function defined on the error-residual η. The parameter s controls the width of the weighting kernel. Associated with the weighting function is an error-kernel which is defined to

$$\rho_s(\eta) = \int_{-\infty}^{\eta} \eta' \Gamma_s(\eta') d\eta' \tag{1}$$

There are many choice of possible weighting functions described in the literature. However, they can be classified according to a broad-based taxonomy based on the derivative $\rho'_s(\eta)$ of the error-kernel. If the derivative is monotonically increasing, then the weighting function is said to be increasing. If the derivative is asymptotically constant, then the weighting function is said to be sigmoidal. Finally, if the derivative asymptotically approaches zero then the weighting function is said to be re-descending.

The standard way to represent the adjacency relations between points is to use the Gaussian proximity matrix. If i and i' are two data points, then the corresponding element of the proximity matrix is given

$$H_D(i, i') = \exp\left[-\frac{1}{2s^2}||\vec{w}_i - \vec{w}_{i'}||^2\right] \tag{2}$$

This weighting function is re-descending. In our previous work [2] we have found that the sigmoidal weighting function, generated by the hyperbolic tangent function

$$H_D(i, i') = \frac{2}{\pi||\vec{w}_i - \vec{w}_{i'}||} \tanh\left[\frac{\pi}{s}||\vec{w}_i - w_{i'}||\right] \tag{3}$$

gives better performance under positional jitter.

2.3 Correspondences

The modal structure of the two point-sets is found by solving the eigenvalue equation $det[H - \lambda I] = 0$ together with the associated eigenvector equation $H\phi_l = \lambda_l \phi_l$, where λ_l is the l^{th} eigenvalue of the matrix H and ϕ_l is the corresponding eigenvector. We order the vectors according to the magnitude of the associated eigenvalues. The ordered column-vectors are used to construct a modal matrix $\Phi = (\phi_1|\phi_2|\phi_3|.....)$. The column index of this matrix refers to the magnitude order of the eigenvalues while the row-index is the index of the original point-set. This modal decomposition is repeated for

both the data and transformed model point-sets to give a data-point modal matrix $\Phi_D = (\phi_1^D|\phi_2^D|\phi_3^D|...|\phi_{|\mathcal{D}|}^D)$ and a model-point modal matrix $\Phi_M = (\phi_1^M|\phi_2^M|\phi_3^M|...|\phi_{|\mathcal{M}|}^M)$. Since the two point-sets are potentially of different size, we truncate the modes of the larger point-set. This corresponds to removing the last $||\mathcal{D}| - |\mathcal{M}||$ rows and columns of the larger matrix. The resulting matrix has $o = \min[\mathcal{D}, \mathcal{M}]$ rows and columns.

The modal matrices can be viewed as inducing a linear transformation on the original identities of the point-sets. Each row of the modal matrix represents one of the original points. The column entries in each row measure how the original point identities are distributed among the different eigen-modes.

Based on this eigendecomposition Shapiro and Brady [11] find correspondences by comparing the rows of the model matrices Φ_M and Φ_D. The decision concerning the correspondences is made on the basis of the similarity of different rows in the modal matrices for the data and the model. The measure of similarity is the Euclidean distance between the elements in the corresponding rows. According to Shapiro and Brady the correspondences are assigned as follows

$$s_{i,j} = \begin{cases} 1 & \text{if } j = \arg\min_{j'} \sum_{l=1}^{o} ||\Phi_D(i,l) - \Phi_M(j',l)||^2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Unfortunately, when the point-sets under study are of different size, i.e. they are subject to structural corruption, then the co-efficients of the modal matrices become unstable and can not be used for correspondence matching. Our aim in to this paper is to suggest a way of overcoming this problem.

3 Hierarchical Correspondence

The coefficients of the modal matrix Φ can be viewed as providing information concerning pairwise clusters of points. Each mode, i.e. each column of the modal matrix Φ, represents a cluster of points. For a given point i the different modal co-efficients $\Phi(i,l)$, $l = 1, ..., |D|$ represent the affinity of the point to the different clusters. The larger the magnitude of the co-efficient, the greater the cluster affinity. In other words, the entries in the columns of the modal matrix represent the membership affinities for the different clusters. The row-entries, on the other hand represent the way in which the individual points are distributed among the different clusters. Here we aim to exploit this property of the modal matrix to develop a fast and robust matching method.

Our idea is based on the simple observation, that while the modal coefficients, i.e. the entries in the columns of the modal matrix, may not be stable under the addition of extra points, the physical centre of the associated cluster will be relatively robust to the addition of outliers. To this end we compute the position of the centre for each cluster. For the mode with eigenvalue λ_l, the position-vector for the cluster centre is

$$\vec{W}_l^D = \frac{\sum_{i=1}^{|D|} |\Phi^D(i,l)| \vec{w}_i}{\sum_{i=1}^{|D|} |\Phi^D(i,l)|} \tag{5}$$

To summarise the arrangement or distribution of the original point-set D, we select the positions of the cluster-centres for the S largest eigenvalues, i.e. the first S columns of Φ. There are a number of ways of choosing S. Here we set the value of S so that the co-efficients of the subsequent columns of Φ are insignificant. If T is a threshold, then the condition is that $|\Phi(i,l)| < T$ for $i = 1, ..., |\mathcal{D}|$ and $l > S$. The number of modes

used might also be set so that there are sufficient points to compute the parameters of the geometric transformation between the two point-sets. In other words 2 modes for the Euclidean transformation and 3 for an affine transformation.

Next, we compute the proximity matrix for these S cluster centres. The elements of the cluster-centre proximity matrix are again computed using the robust weighting kernel and are given by

$$G_D(l,l') = \frac{2}{\pi||\vec{W}_l^D - \vec{W}_{l'}^D||}\tanh\left[\frac{\pi}{s}||\vec{W}_l^D - \vec{W}_{l'}^D||\right] \tag{6}$$

Our idea is to use the modes of the $S \times S$ cluster-centre proximity matrix for the purposes of matching. Accordingly, we solve the equation $det[U_D - \Lambda I] = 0$ to locate the eigenvalues Λ of G_D and then compute the associated eigenvectors F_Λ^D by solving the equation $G_D F_{\Lambda_L}^D = \Lambda_L^D F_{\Lambda_L}^D$. The eigenvectors are used as the columns of the cluster-centre modal matrix $U_D = (F_{\Lambda_1}^D|F_{\Lambda_2}^D|.....|F_{\Lambda_S}^D)$. This procedure is also repeated for the first S modes of the point-set M. The resulting $S \times S$ modal matrix is denoted by U_M.

In the next section we will describe an iterative method which locates correspondences by aligning the cluster centers of the data point-set with those of the model point-set. This is achieved using an affine transformation. At iteration n of the algorithm, the co-ordinate vector of the ith cluster-center of the aligned set of points is $\vec{W}_i^{(n)}$. The modal matrix for these aligned cluster-centres is denoted by $U_D^{(n)}$.

We use the elements of the two modal matrices to compute the probabilities of correspondence match. This is done by comparing the elements of the two matrices on a row-by-row basis. A simple way of computing the probabilities is to assume that the vectors are subject to Gaussian measurement errors. The shortcoming of this method for computing the correspondence probabilities is the effect of outlier measurement errors on the individual components of the eigen-vectors. When there is a significant difference between one or more of the components of the eigenvectors, then these errors dominate the argument of the exponentials. This will have the tendency to flatten the distribution and will result in ambiguity and equivocation concerning the pattern of correspondences. One way to make the computation of correspondences robust to outlier measurement error is to accumulate probability on a component by component basis over the eigenvectors. To do this we define the correspondence probability to be

$$\zeta_{i,j}^{(n)} = \frac{\sum_{L=1}^{S}\exp\left[-\mu||U_D^{(n)}(i,L) - U_M(j,L)||^2\right]}{\sum_{j=1}^{S}\sum_{L=1}^{S}\exp\left[-\mu||U_D^{(n)}(i,L) - U_M(j,L)||^2\right]} \tag{7}$$

where $U_D^{(n)}$ is the modal matrix computed from the transformed point positions and μ is a constant. In this way large measurement errors contribute insignificantly through the individual exponentials appearing under the summation over the components of the eigenvectors.

4 Dual-step EM Algorithm

Once the two cluster-centre modal matrices are to hand, then we can attempt to locate correspondences between the clusters in the two point-sets. Here we use the dual-step EM algorithm of Cross and Hancock [3] to align the cluster-centres. Cross and Hancock's contribution was to present an extension of the standard EM algorithm [4] in which the

structural consistency of correspondences matches can be used to gate contributions to the expected log-likelihood function [3]. This idea is closely related to the hierarchical mixture of experts algorithm of Jordan and Jacobs [7]. However, the method uses a dictionary method for computing the correspondence probabilities which is both localised and time consuming. The aim here is to replace the dictionary-based method used to compute the probabilities with a hierarchical spectral method.

4.1 Affine Geometry

Suppose that the matrix A represents the geometric transformation that best aligns the set of cluster centres in the data and the model. In this paper we confine our attention to affine transformations. The affine transformation has six free parameters. These model the two components of translation of the origin on the image plane, the overall rotation of the co-ordinate system, the overall scale together with the two parameters of shear. These parameters can be combined succinctly into an augmented matrix that takes the form

$$A^{(n)} = \begin{pmatrix} a_{1,1}^{(n)} & a_{1,2}^{(n)} & a_{1,3}^{(n)} \\ a_{2,1}^{(n)} & a_{2,2}^{(n)} & a_{2,3}^{(n)} \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

With this representation, the affine transformation of the data point-set cluster centres is computed using the matrix multiplication $\vec{W}_i^{(n)} = A^{(n)}\vec{W}_i$. The superscript n indicates that the parameters are taken from the n^{th} iteration of our algorithm.

4.2 Expected Log-likelihood

According to Cross and Hancock we seek both correspondence matches (i.e. the function f) and transformation parameters which maximise the expected log-likelihood

$$Q(A^{(n+1)}|A^{(n)}) = \sum_{i=1}^{S}\sum_{j=1}^{S} P(\vec{W}_j^M|\vec{W}_i^D, A^{(n)}) \times \zeta_{i,j}^{(n)} \ln p(\vec{W}_i^D|\vec{W}_j^M, A^{(n)}). \tag{9}$$

The meaning of this expected log-likelihood function requires further comment. The measurement densities $p(\vec{W}_i^D|\vec{W}_j^M, A^{(n+1)})$ model the distribution of error-residuals between the data point-set cluster centre $\vec{W}_i^D$ and the the model points-set cluster centre $\vec{W}_j^M$ at iteration n of the algorithm. The log-likelihood contributions at iteration $n+1$ are weighted by the *a posteriori* measurement probabilities $P(\vec{W}_j^M|\vec{W}_i^D, A^{(n)})$ computed at the previous iteration n of the algorithm. The individual contributions to the expected log-likelihood function are gated by the correspondence matching probabilities $\zeta_{i,j}^{(n)}$.

4.3 Expectation

In the expectation step of the EM algorithm the *a posteriori* measurement probabilities are updated by substituting the updated cluster centre position vectors into the conditional measurement distribution. Using the Bayes rule, we can re-write the *a posteriori* measurement probabilities in terms of the components of the corresponding conditional measurement densities

$$P(\vec{W}_j^M|\vec{W}_i^D, A^{(n)}) = \frac{\alpha_j^{(n)} p(\vec{W}_i^D|\vec{W}_j^M, A^{(n)})}{\sum_{j'=1}^{S} \alpha_{j'}^{(n)} p(\vec{W}_i^D|\vec{W}_{j'}^M, A^{(n)})}. \tag{10}$$

The mixing proportions are computed by averaging the *a posteriori* probabilities over the set of cluster-centres for the data-points, i.e.

$$\alpha_j^{(n+1)} = \frac{1}{S}\sum_{i=1}^{S} P(\vec{W}_j^M|\vec{W}_i^D, A^{(n)})$$

In order to proceed with the development of a point registration process we require a model for the conditional measurement densities. Here we assume that the required model can be specified in terms of a multivariate Gaussian distribution. Accordingly we write

$$p(\vec{W}_i^D|\vec{W}_j^M, A^{(n)}) = \frac{1}{2\pi\sqrt{|\Sigma|}}\exp\left[-\frac{1}{2}\vec{\varepsilon}_{ij}^{(n)\top}\Sigma^{-1}\vec{\varepsilon}_{ij}^{(n)}\right]. \tag{11}$$

In the above expression Σ is the variance-covariance matrix for the position errors and $\vec{\varepsilon}_{ij}^{(n)} = \vec{W}_j^M - A^{(n)}\vec{W}_i^D$.

4.4 Maximisation

The dual step EM algorithm iterates between the two interleaved maximisation steps for alignment parameter estimation and estimating correspondence assignments.

Point correspondences are sought so as to maximise the *a posteriori* probability of structural match. The update formula is

$$s_{i,j}^{n)} = \begin{cases} 1 & \text{if } j = \arg\max_{j=1}^{S} P(\vec{W}_j^M|\vec{W}_i^D, A^{(n)})\zeta_{i,j}^{(n)} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

In the case of affine geometry, the transformation is linear in the parameters. This allows us to locate the maximum-likelihood parameters directly by solving a system of saddle-point equations for the independent affine parameters. The solution matrix is given by

$$A^{(n+1)} = \left[\sum_{i=1}^{S}\sum_{j=1}^{S} P_{i,j}^{(n)}\zeta_{i,j}^{(n)}\vec{W}_i^D\, R^T\, \vec{W}_i^{DT}\,\Sigma^{-1}\right]^{-1} \times \left[\sum_{i=1}^{S}\sum_{j=1}^{S} P_{i,j}^{(n)}\zeta_{i,j}^{(n)}\vec{W}_j^M\, R^T\, \vec{W}_i^{DT}\,\Sigma^{-1}\right] \tag{13}$$

where

$$R = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{14}$$

is a projection matrix and we use the shorthand $P_{i,j}^{(n)} = P(\vec{W}_j^M|\vec{W}_i^D, A^{(n)})$ for the a posteriori measurement probabilities.

5 Experiments

In this section we investigate the performance of the different methods of spectral correspondence reported in this paper. We investigate the noise sensitivity of the method using synthetic point sets. The random point-sets used in our experiments have been generated as follows. In order to synthesise point-sets which have a well defined cluster structure, we generate seed points at random image locations. We then generate a distribution of points about each seed whose displacements from the seed-point are sampled from a two dimensional Gaussian distribution of zero mean and known variance. The simulate the

effects of noise we have added two types of random error to the pont-sets generated in this way. Firstly, we have added Gaussian point-position error or jitter to the point locations. Secondly, we have added contaminating points at random image locations. We use $S = 7$ modes to compute the affine transformation between the point-sets. The recovered transformation parameters are used to align the two point-sets and the number of closest matches that are in correct correspondence are counted.

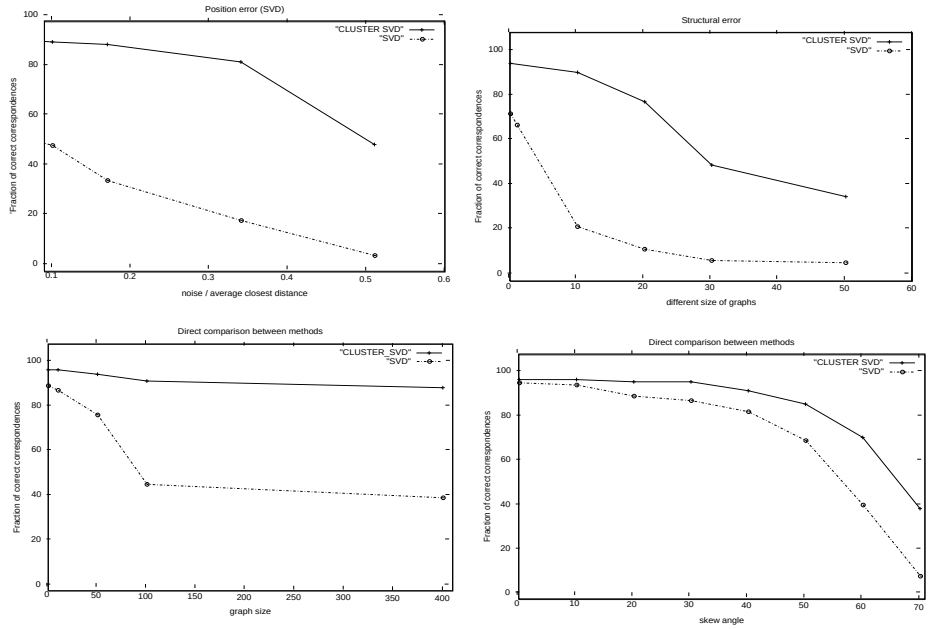

Figure 1: Experimental results

In Figure 1 we compare our new correspondence method (solid curve) with that of Shapiro and Brady (dotted curve). Figure 1a investigates the effect of added point contamination. The plot shows the fraction of correct correspondences for a point-set of size 100 as a function of the number of added points, i.e. the size difference of the two point-sets. The Shapiro and Brady method fails rapidly as soon as the size difference becomes greater than 5%. By contrast the hierarchical method maintains much better correspondences. In Figure 1b we investigate the effect of measurement error or positional jitter. The plot shows the fraction of correct correspondences a function of the ratio of the standard deviation of the added Gaussian noise to the average inter-point distance. The hierarchical method consistently outperforms that of Shapiro and Brady by a margin of some 35%. Next, Figure 1c shows the fraction of correct correspondences as a function of the size of the point sets. Here the standard deviation of the added Gaussian noise is kept fixed and the size of the point-sets is increased. The two point sets are of the same size. There is no addition or deletion of points. The size of the point-sets is increased by adding new points at random positions. As a result the point-density increases. The main effects to note are that the Shapiro and Brady method degrades rapidly with point-density while the performance of our new method is almost unaffected. This is an important observation since it shows that our method is not sensitive to the dilution or merging of the cluster structure due to in-fill by additional points. Finally, Figure 1d shows the fraction

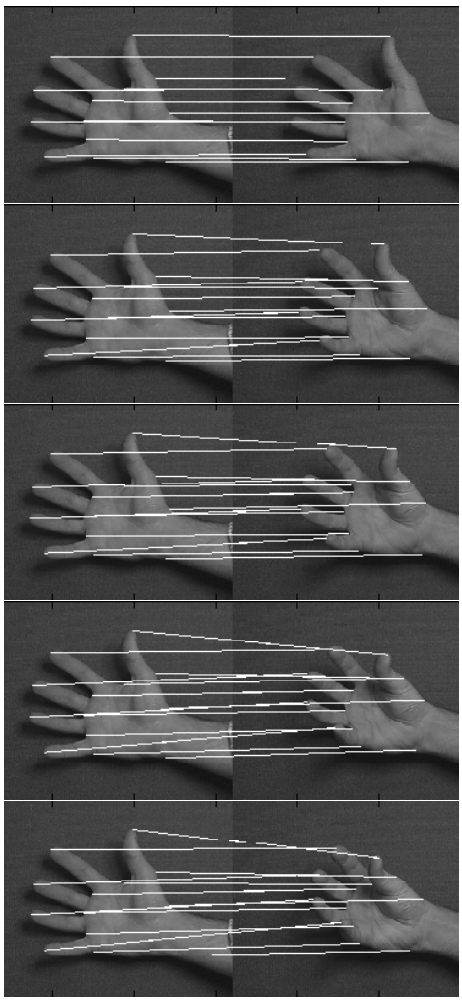

Figure 2: Matches-spectral clusters

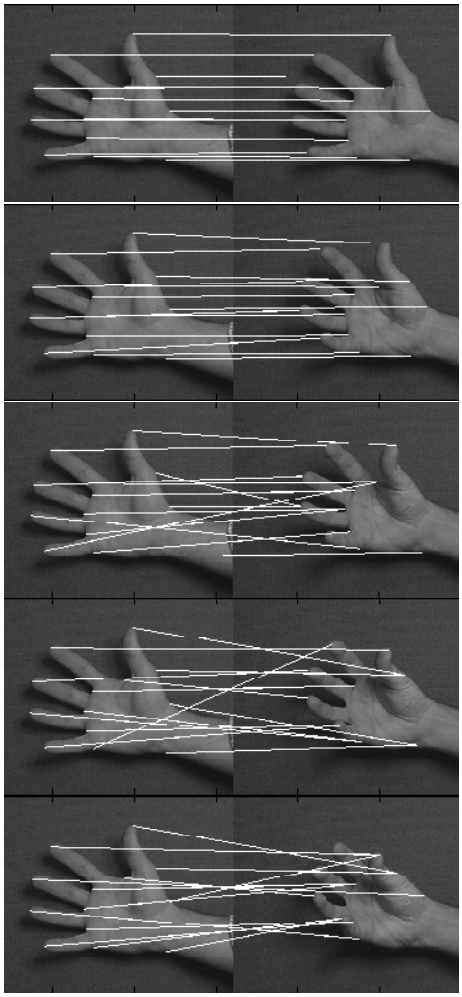

Figure 3: Matches- Shapiro and Brady

of correct correspondences as a function of the affine skew angle. Again, our new method performs better under severe affine distortion.

We have experimented with our new correspondence method on a motion sequence. This shows a hand being gradually clenched to form a fist. Figures 2 and 3 respectively show the sequence of correspondnece matches obtained with the hierarchical clustering method reported in this paper and the method of Shapiro and Brady. In each panel the image on the left-hand side is the initial image in the sequence. The images in the right-hand panel are those obtained after 1, 10, 15, 20 and 25 frames have elapsed. Here the sequence is captured at a rate of 10 frames per second. The matches shown are directly from the left-hand frame to the right-hand frame, i.e. no intermediate frames are used. In the case of the clustering method, the matches are all correct in each frame. In the case of the Shapiro and Brady algorithm failure occurs after 10 frames.

6 Conclusions

This paper has focussed on how to improve the robustness of the Shapiro and Brady method of modal or spectral correspondence to the effects of clutter and dropout. Our approach is a hierarchical one. We commence from the proximity matrix for the raw point-sets. We use the co-eeficients of the modal matrix of the initial proximity matrix to compute the centrepoints of modal clusters. These cluster centres, themselves, are used to compute a second proximity matrix. By computing correspondence probabilities using the modal co-efficients of this second proximity matrix, we are able to estimate accurate affine transformation parameters. In this way, we can improve the robustness of spectral correspondence. Although the method reported here has been demonstrated on point-patterns it can easily be extended to other proximity based representations such as grey-scale difference [12] or line-attributes [9].

References

[1] F.R.K. Chung, *Spectral Graph Theory*, CBMS series **92**, AMS Ed., 1997.

[2] M. Carcassoni and E.R Hancock, "Point Pattern Matching with Robust Spectral Correspondence", *CVPR 2000*.

[3] A.D.J. Cross and E.R. Hancock, "Graph matching with a dual step EM algorithm", *IEEE PAMI*, **20**, pp. 1236–1253, 1998.

[4] A.P. Dempster, Laird N.M. and Rubin D.B., "Maximum-likelihood from incomplete data via the EM algorithm", *J. Royal Statistical Soc. Ser. B (methodological)*, **39**, pp. 1-38, 1977.

[5] R. Horaud and H. Sossa, "Polyhedral Object Recognition by Indexing", *Pattern Recognition*, **28**, pp. 1855-1870, 1995.

[6] K.Inoue and K. Urahama, "Sequential fuzzy cluster extraction by a graph spectral method", *Pattern Recognition Letters*, **20**, pp. 699-705, 1999.

[7] M.I. Jordan and R.A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm", *Neural Computation*, **6**, pp. 181-214, 1994.

[8] G.L. Scott and H.C. Longuet-Higgins, "An algorithm for associating the features of 2 images", *Proceedings of the Royal Society of London Series B (Biological)*, **244**, pp. 21–26, 1991.

[9] K. Sengupta and K.L.Boyer, "Modelbase partitioning using property matrix spectra", *Computer Vision and Image Understanding*, **70**:2, pp. 177-196, 1998.

[10] A. Shokoufandeh, S.J. Dickinson, K. Siddiqi and S.W. Zucker, "Indexing using a spectral encoding of topological structure", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.491-497, 1999.

[11] L.S. Shapiro and J.M. Brady, "Feature-based correspondence - an eigenvector approach", *Image and Vision Computing*, **10**, pp. 283–288, 1992.

[12] J. Shi and J.Malik, "Normalized cuts and image segmentation", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[13] S. Umeyama, "An eigen decomposition approach to weighted graph matching problems", *IEEE PAMI*, **10**, pp. 695–703, 1988.